Multi-DMA Virtualization within Virtualized PCIe® Systems

Stephane Hauradou

Vice President

PLDA

Disclaimer

<u>Presentation Disclaimer</u>: All opinions, judgments, recommendations, etc. that are presented herein are the opinions of the presenter of the material and do not necessarily reflect the opinions of the PCI-SIG®.

Agenda

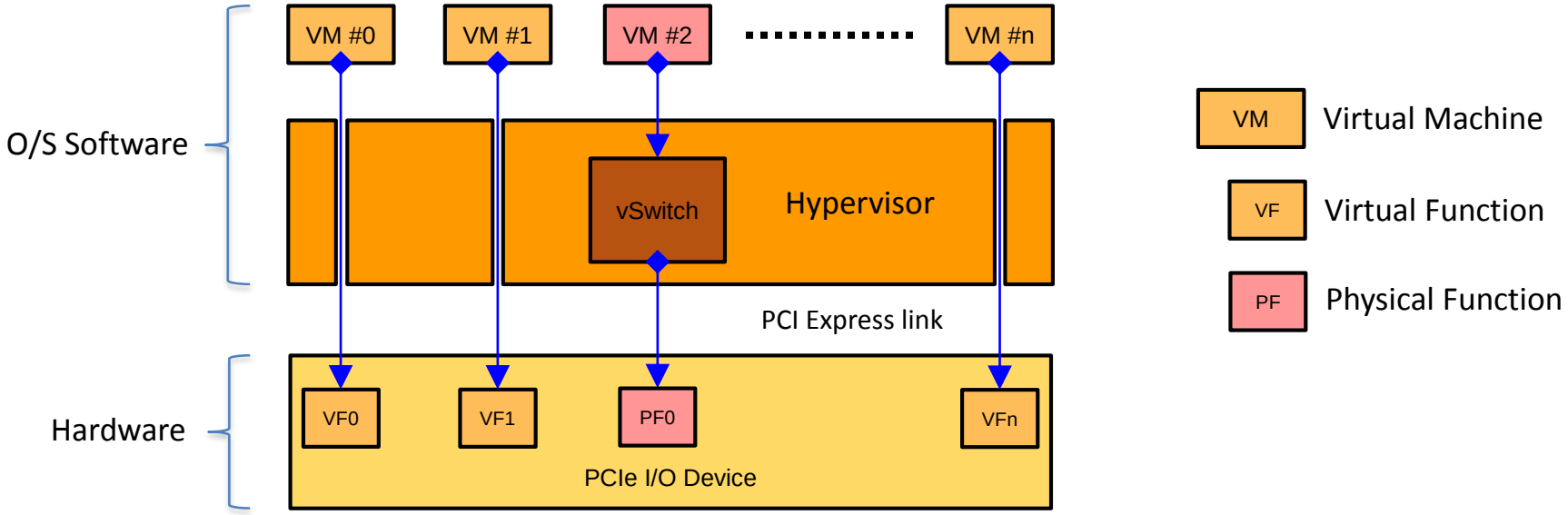
- **Data Center Virtualization and the need from efficient DMA**
- **Traditional DMA approach**
  - Architecture, limitations
- **Proposed Virtualized DMA**
  - Requirements, architecture, performance, challenges
- **The vDMA IP**
  - Features, availability, roadmap
- **Q&A**

Virtualization in the Data Center

- **Exponential traffic increase**
- **Move to SDx, virtualized hardware**
- **Network, Storage, Compute convergence**

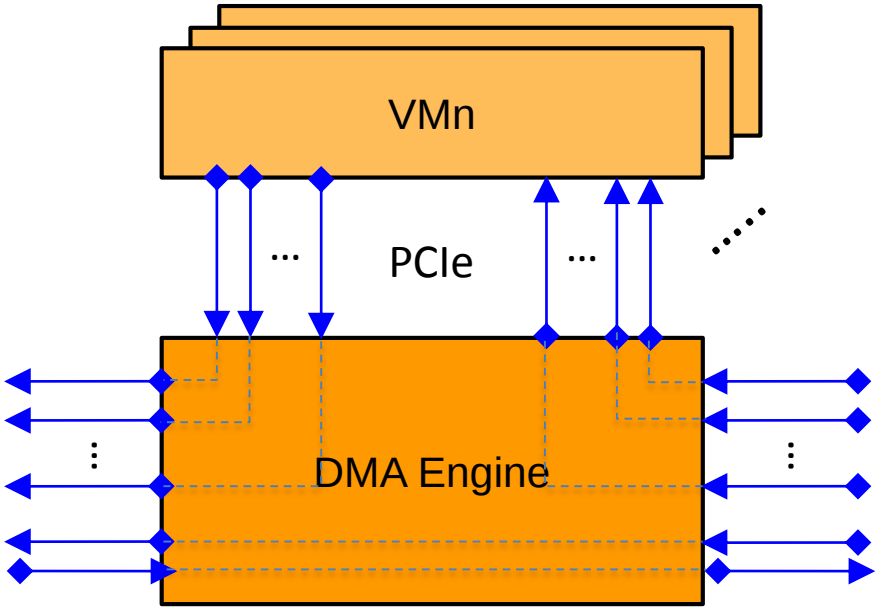
About PCIe® Virtualization

- **PCI-SIG PCIe® Specification Extension**
  - SR-IOV, MR-IOV
- **Allows multiple VMs to share an I/O device**
  - Bare metal performance through Hypervisor bypass



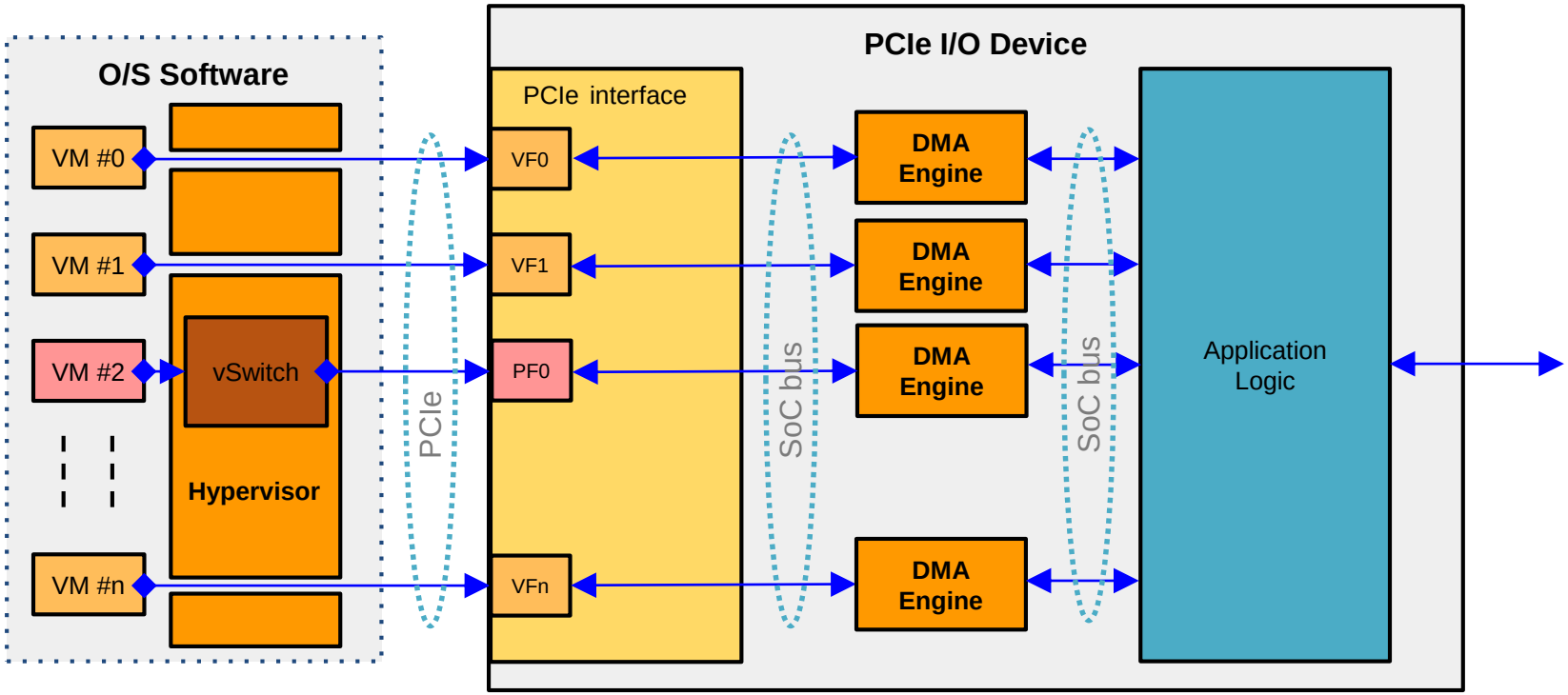
VMs Need Efficient DMA

- **Multiple channels per VM, any direction**
- **Fair bandwidth sharing between VM**
- **Non-blocking, isolated VMs**

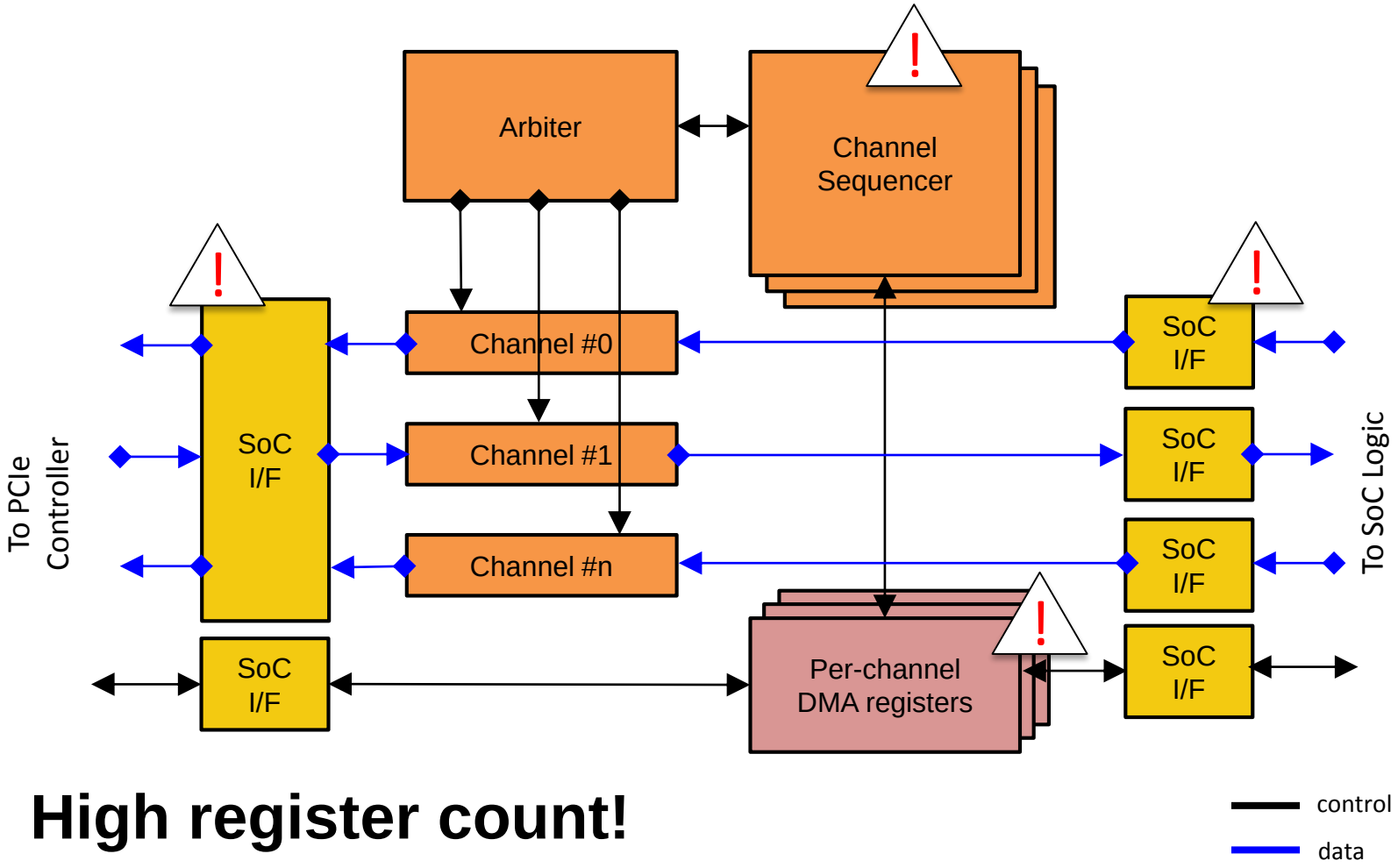


In an Ideal World…

- **Each VM has its own DMA resources**
- **N-channel DMA per VF**



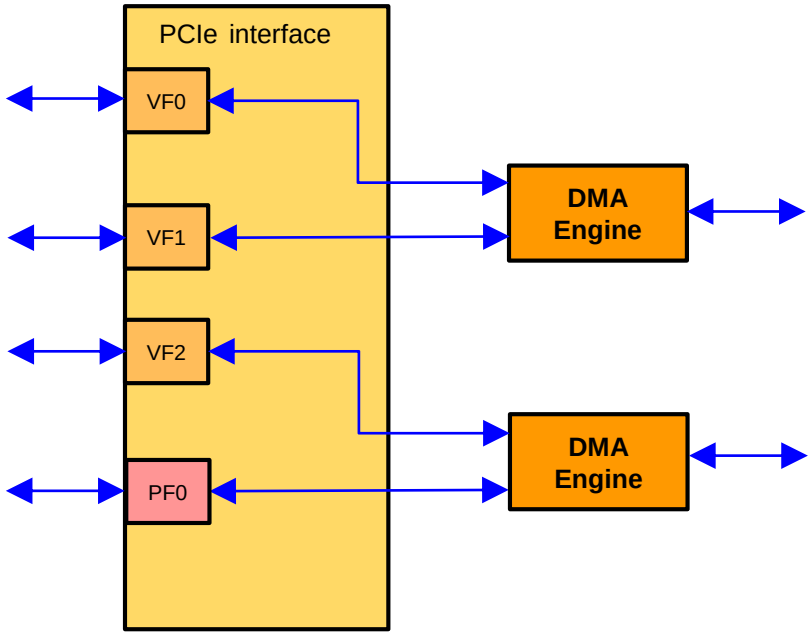
Traditional DMA Architecture



- **High register count!**

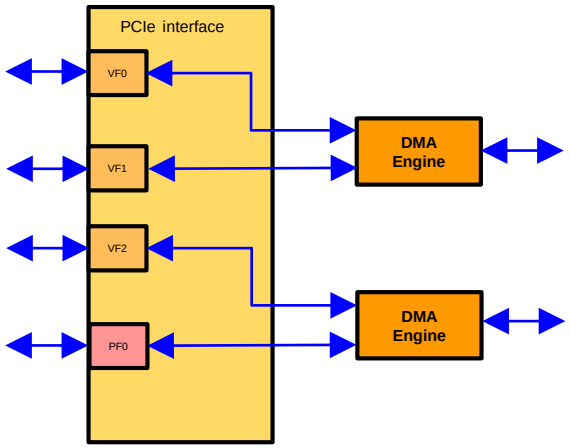
Traditional IOV DMA Approach

- **DMA sharing**



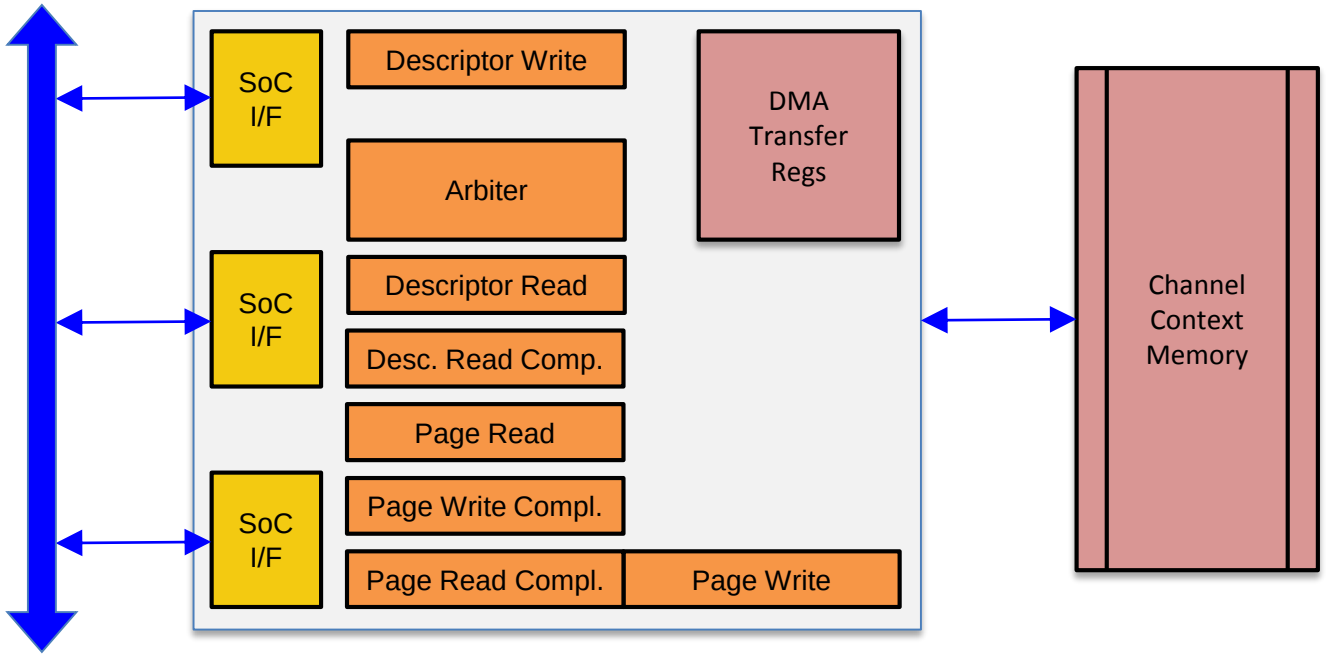
IOV DMA Limitations

- **VF context switching latency**
- **VF blocking**
- **VM isolation**
- **Scalability**



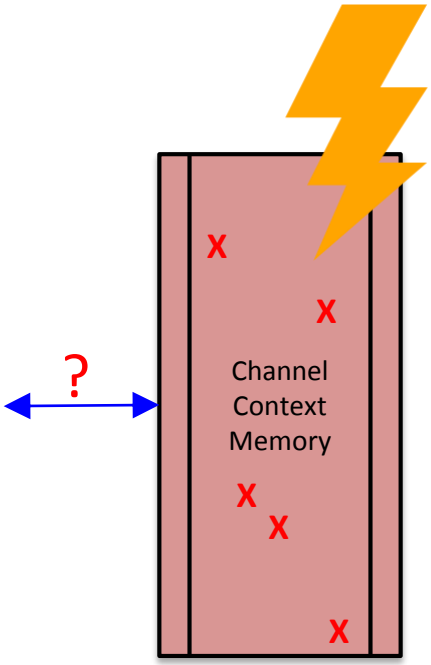
A Novel Approach to DMA

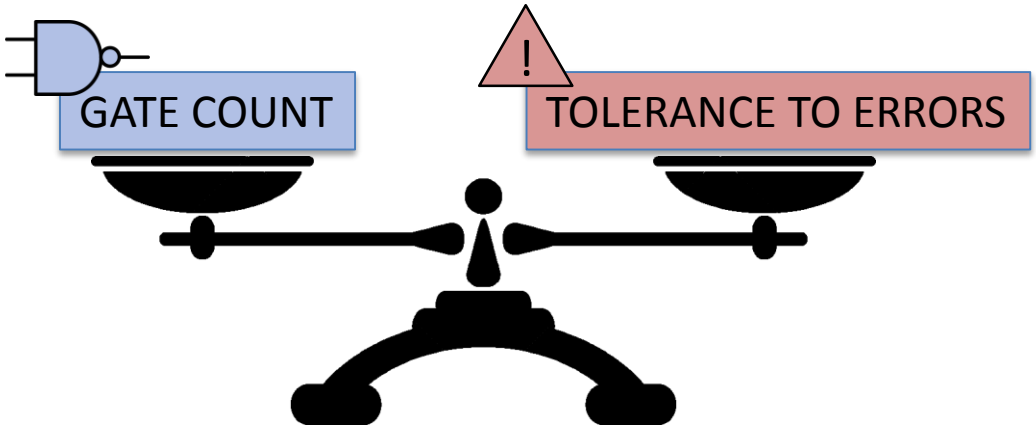
- **Concurrent DMA tasks**
- **External Context memory**
- **Channel data multiplexing**



Context Memory Considerations

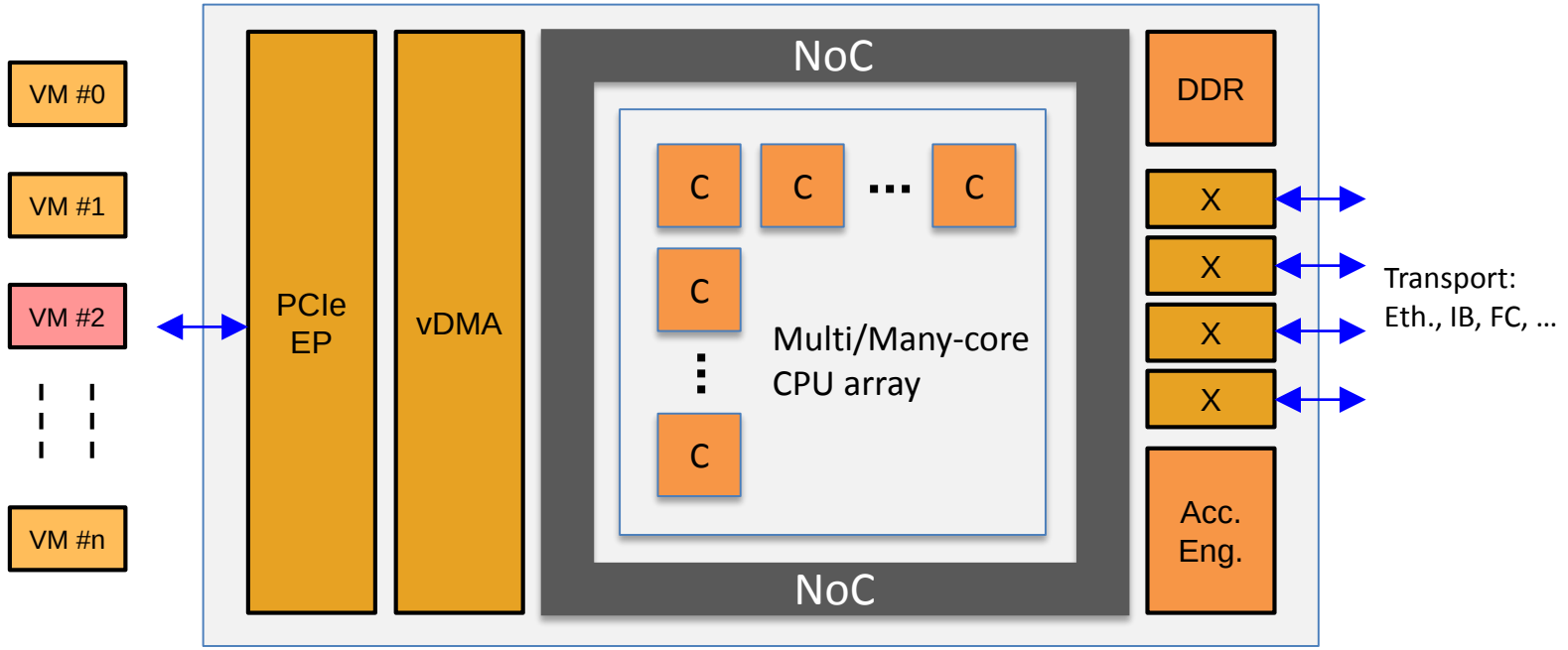
- **Parity, checksum, CRC**
  - Report errors to VM
- **ECC**
  - Only report NC errors to VM





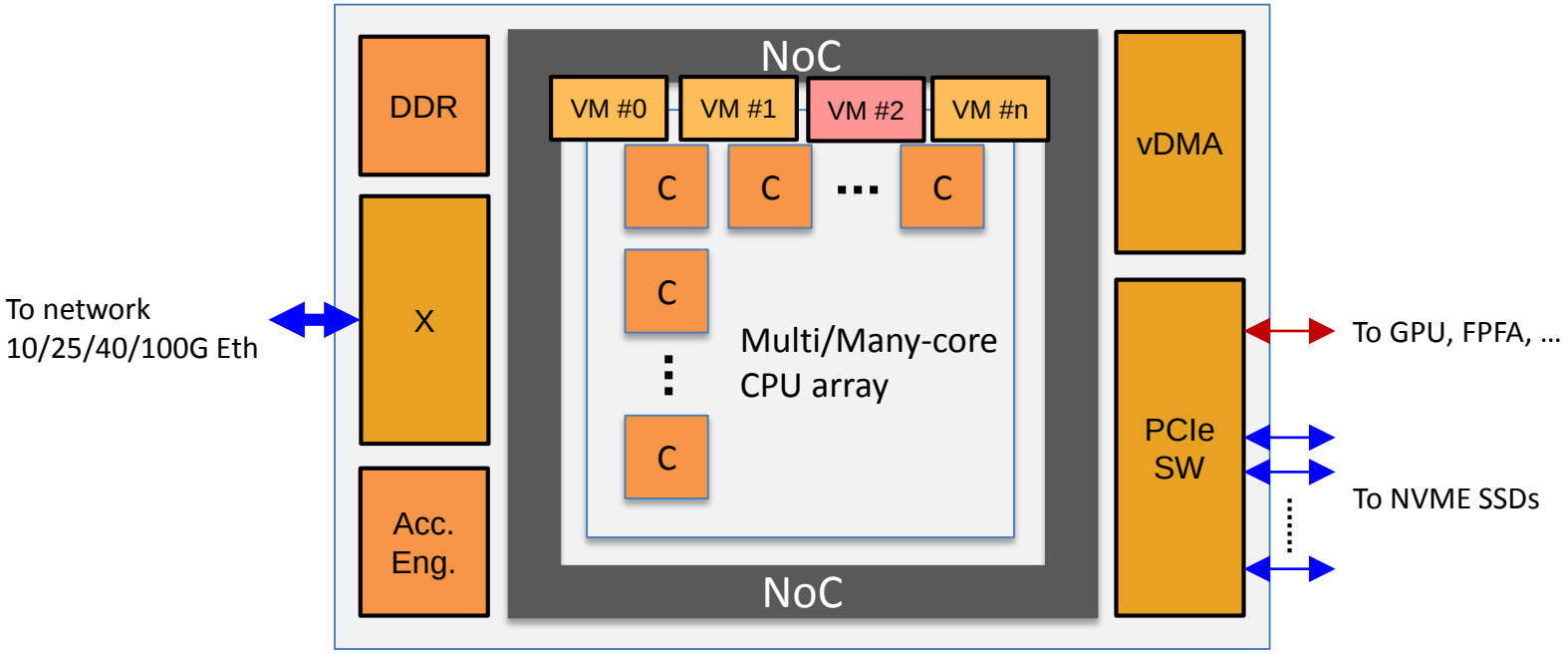
Virtualized DMA Use Models (1)

- **Data Center SoCs for Network & Storage offload**
  - Flow processors, SmartNICs, Storage controllers



Virtualized DMA Use Models (2)

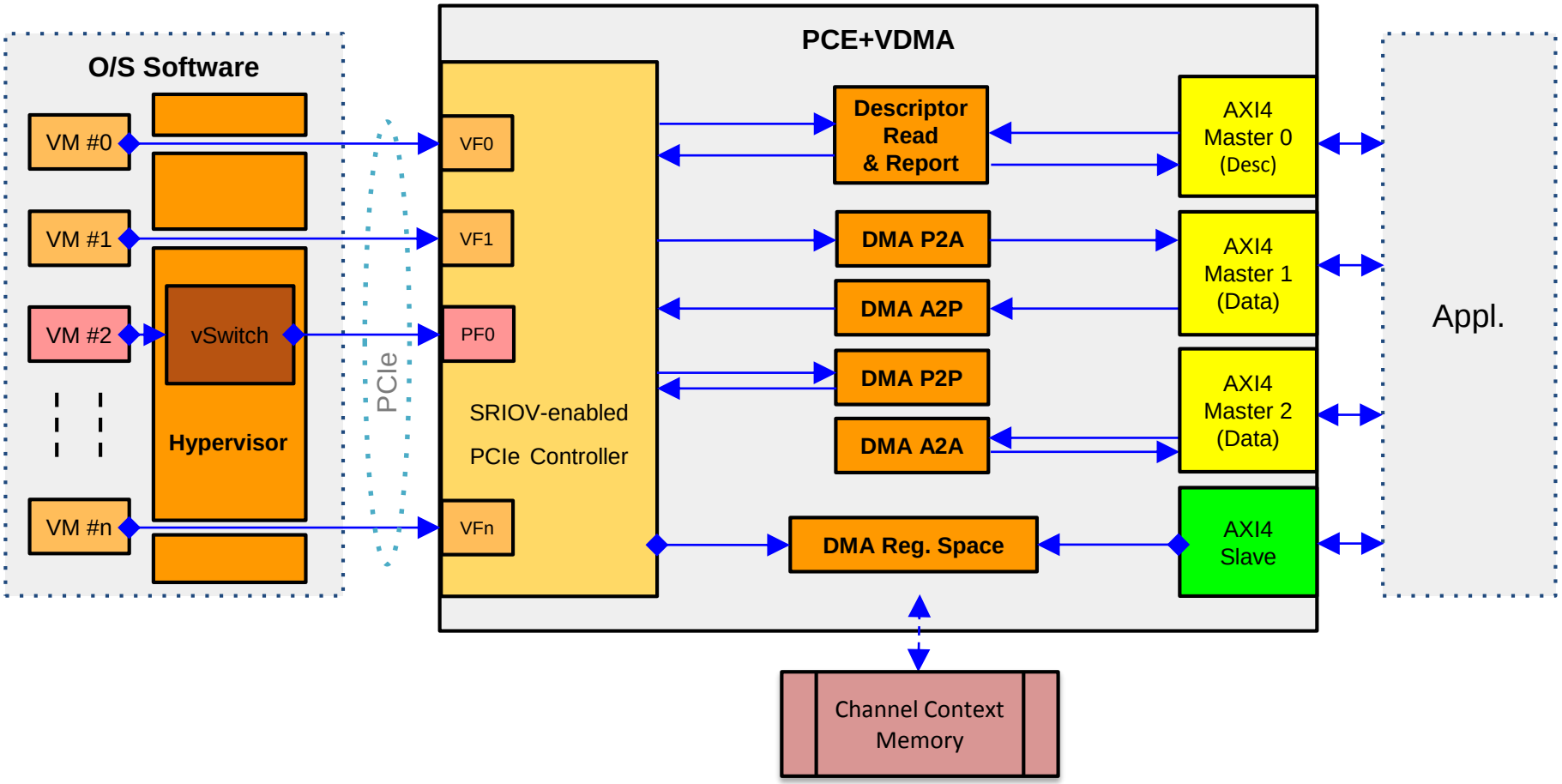
- **Data Center SoCs for Network & Storage offload**
  - NVMe Storage controllers



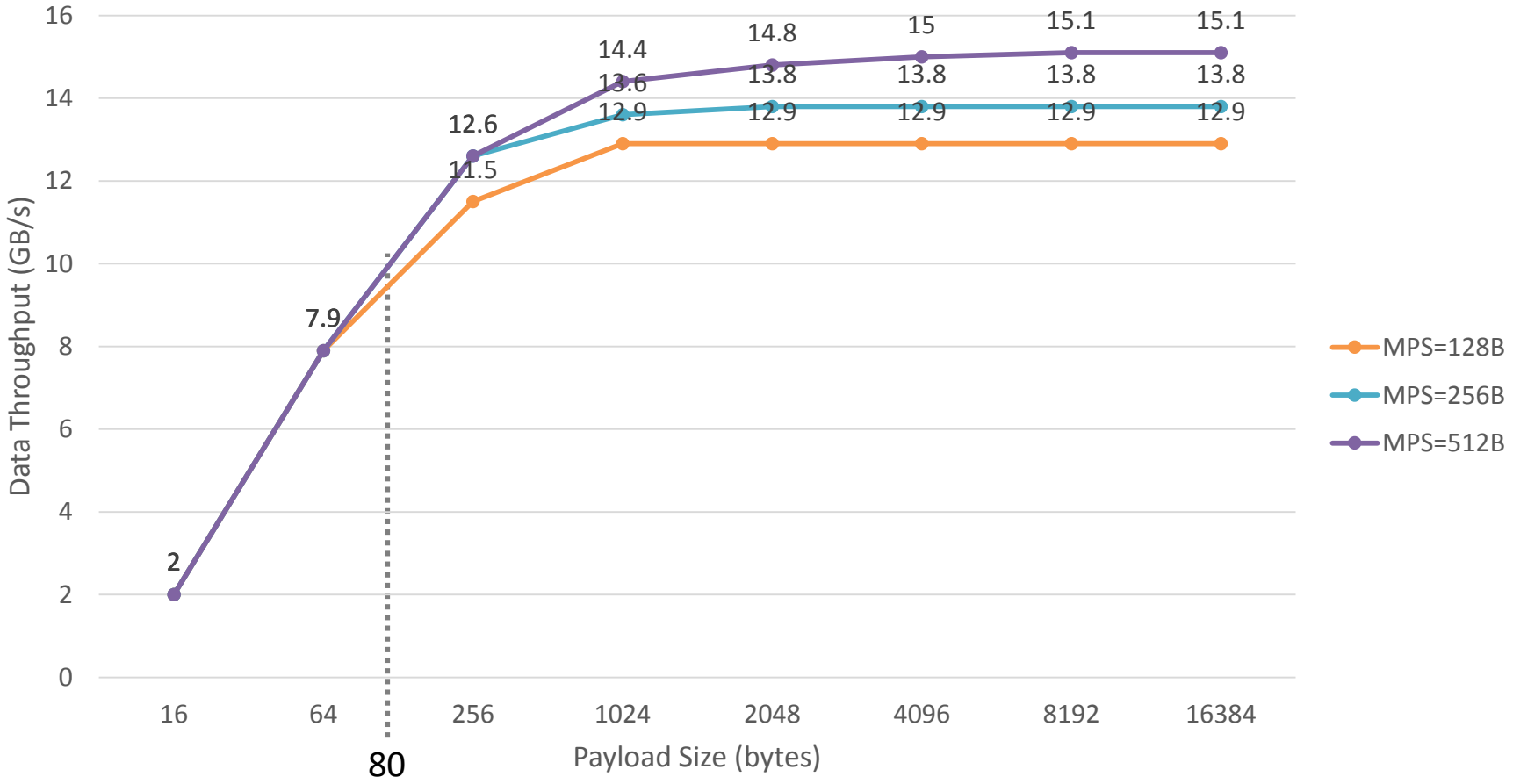
Virtualized DMA Requirements

- **Performance**
  - High throughput, small packets
  - Any # of DMA channel per VM
  - Non-blocking operations
- **Low Latency**
  - VM switching
  - QoS for VMs
- **Security**
  - VM isolation

Virtualized PCIe DMA

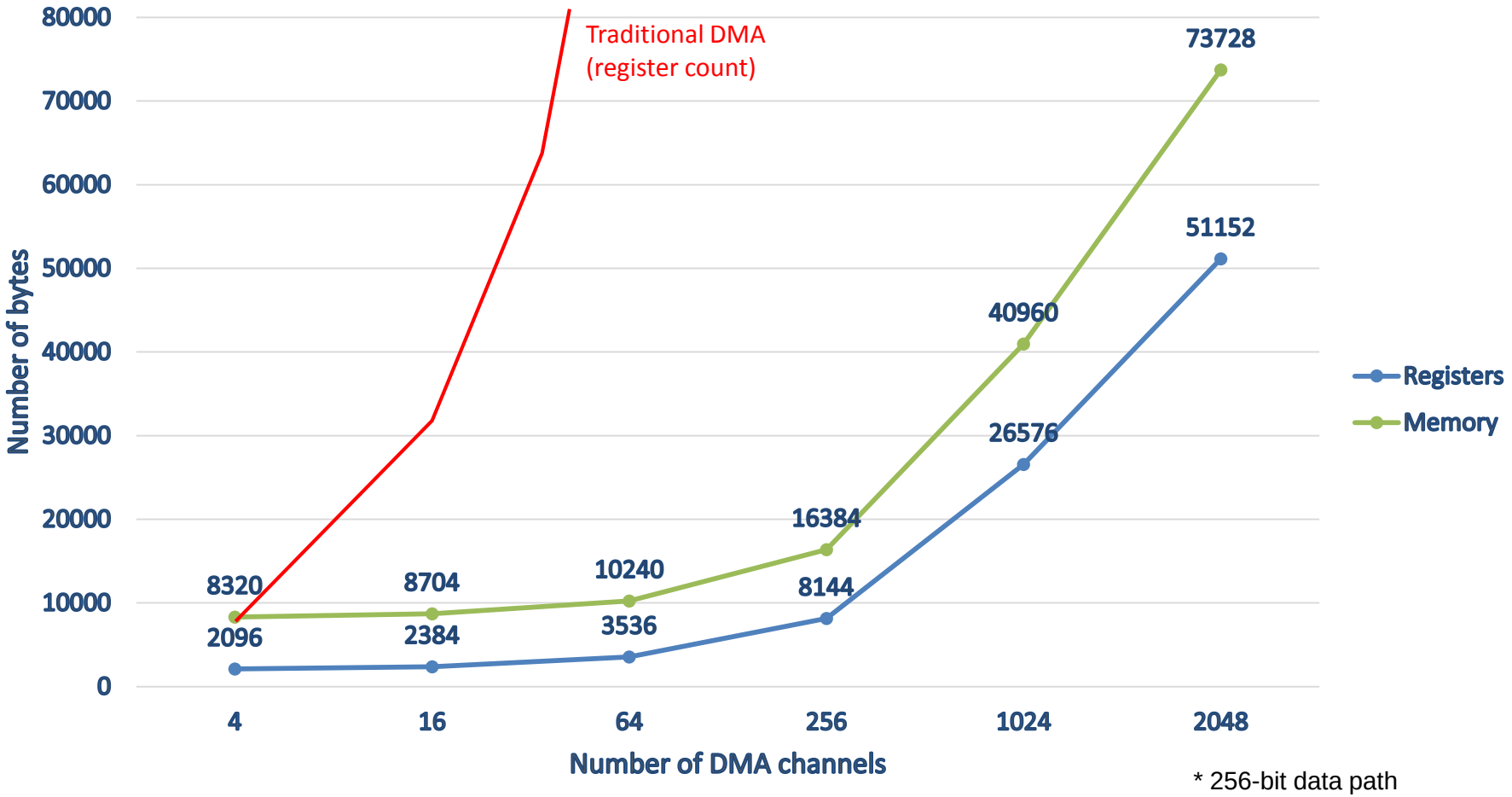


vDMA Performance



* PCIe x16 8GT/s
* 128 Outstanding Read Requests

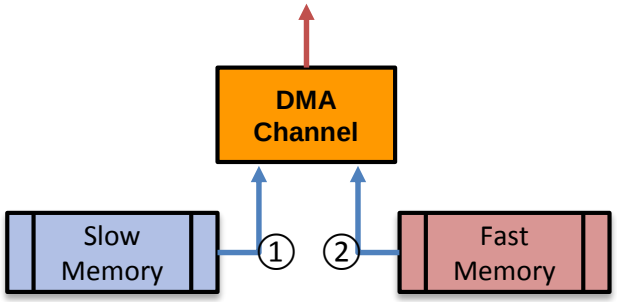
vDMA Resources



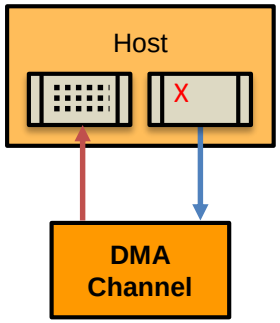
* 256-bit data path
* Excludes PCIe controller logic

Technical Implementation
Non-Blocking Operation

Channel latency mismatch

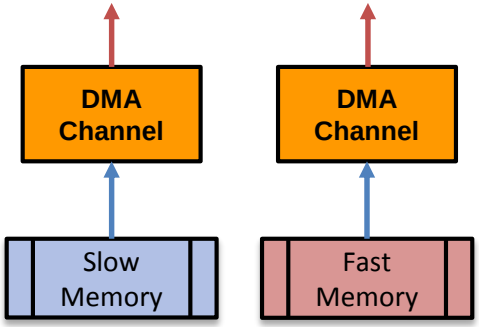


Host buffer/data unavailable

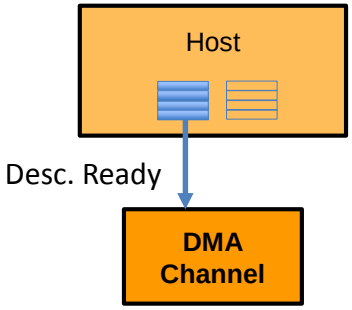


VM crash

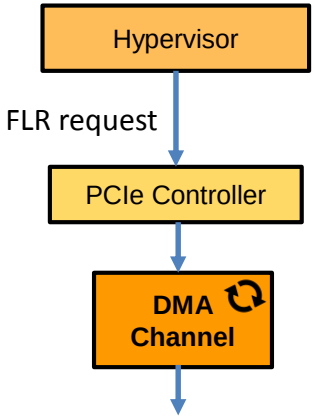
Separate channels



Host-to-DMA protocol

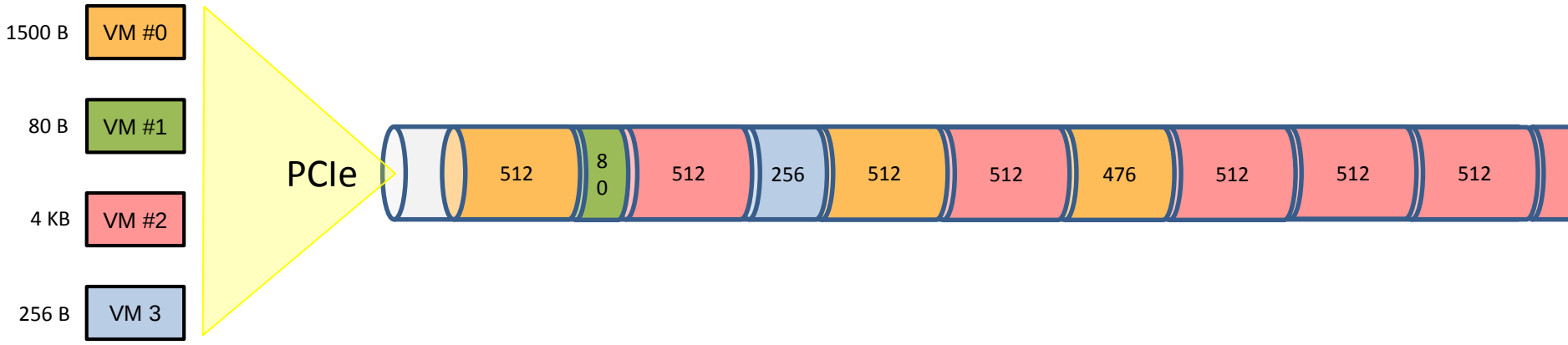


Channel reset via FLR



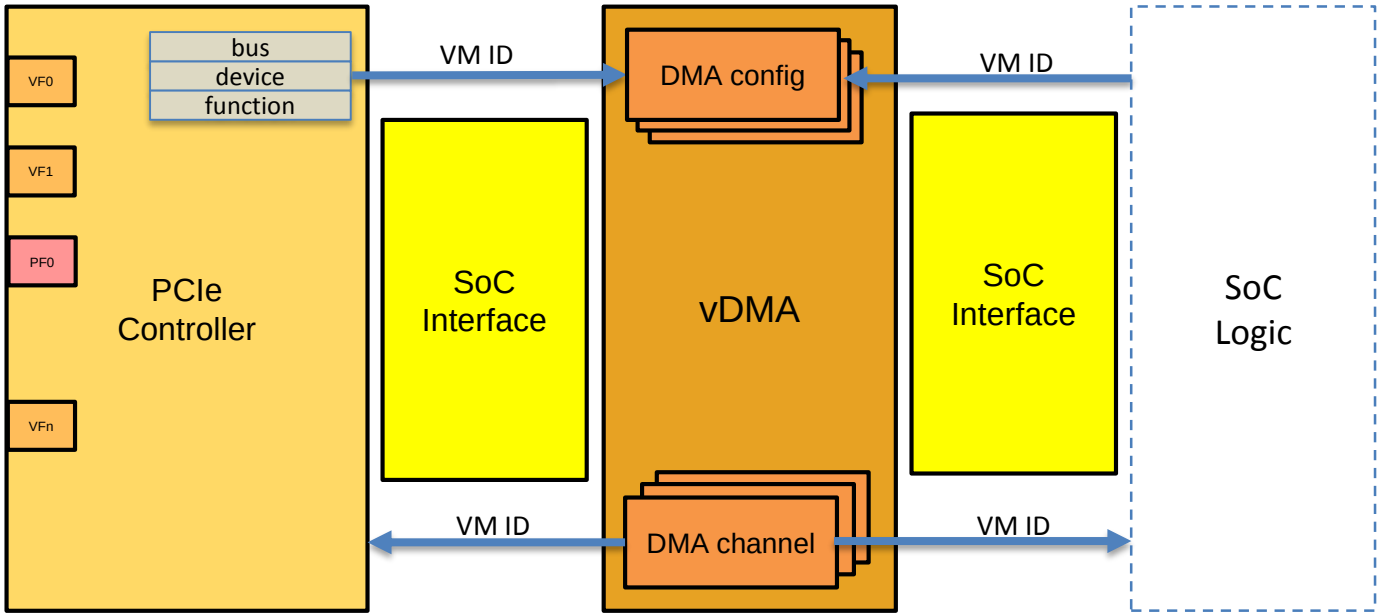
Technical Implementation
Fair Arbitration Between VMs

- **Round-robin bandwidth sharing**
  - Statically configurable
  - Active channels only



Technical Implementation
VM Isolation

- **Use sideband signaling**



Technical Implementation 10GB/s for 80b Packets

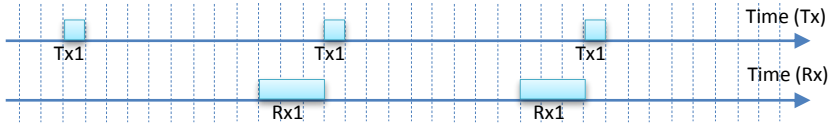


Fig. 1 - 1 outstanding request, small packets

$$\mathbf{BW} = \frac{\boldsymbol{S} * \boldsymbol{N}}{\boldsymbol{RTL}}$$

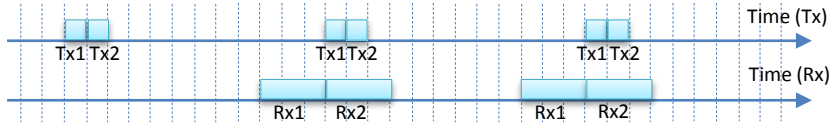
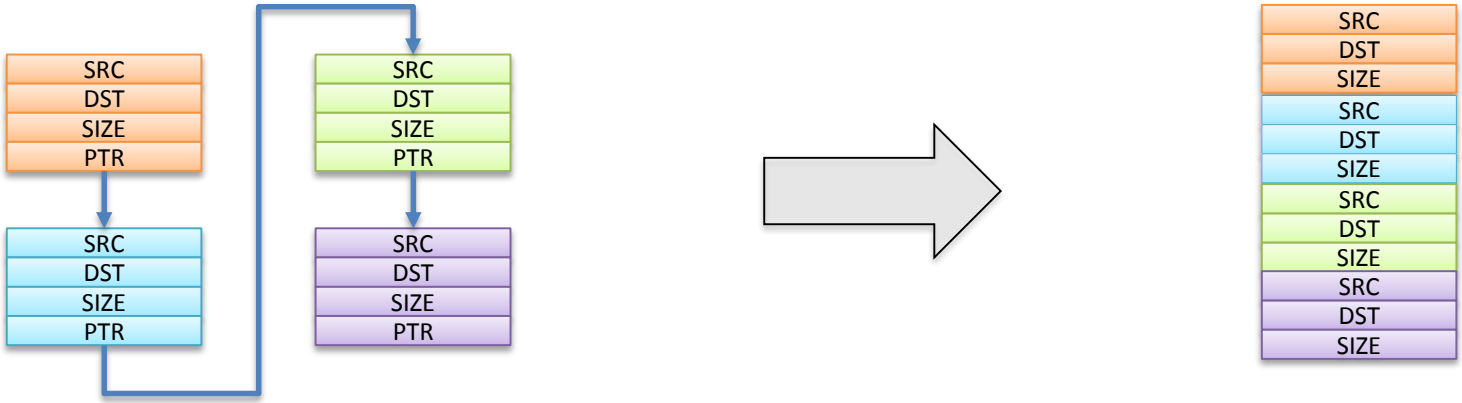


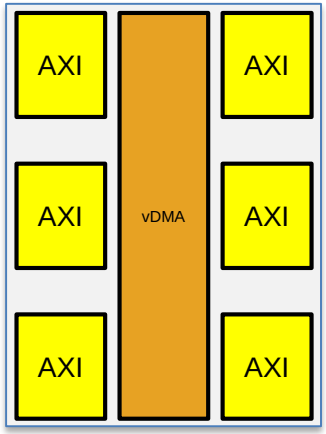
Fig. 2 - 2 outstanding requests, small packet

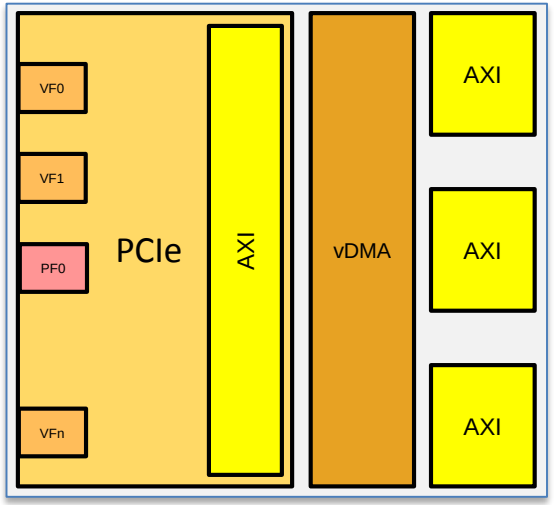
- **Increase N**



- **Enforce contiguous DMA descriptors**

The vDMA IP





- **Standalone**
  - PCIe agnostic
- **Integrated PCIe**
  - Xilinx/Intel PCIe Hard IP
  - PLDA XpressRICH Soft IP
- **2048+ DMA channels**
- **256-bit data path**
  - PCIe 8GT/s x8 and 16GT/s x4
- **Availability**
  - Aug. 2017

Future Improvements

- **QoS (Traffic Policies)**
  - VM priority (ex. Weighted Round Robin)
- **512-bit architecture**
  - 16GT/s x8 and 8GT/s x16 throughput
- **Data integrity**
  - CRC, parity, ECC
- **Streaming**
  - AXI streams

Thank you for attending the PCI-SIG Developers Conference 2017.

For more information please go to www.pcisig.com